Content-Aware Optimization of Tiled 360° Video Streaming Over Cellular Network

Saurabh S. Dange, Shashwat Kumar, and Dr. Antony Franklin A.

NeWS Lab
Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

IEEE WiMob 2021, Virtual

https://newslab.iith.ac.in/

Contents

Introduction

- 360° videos, also known as immersive videos or spherical videos, are video recordings where a view in every direction is recorded at the same time.
- 360° video playback provides 3-degree of freedom.

Fig. 360° Video EQR Projection.

Introduction

- Field of View (FoV) is an area observable by the user at any instance.
- In comparison to a planar 4K video, a 360° videos may require up to 10x bandwidth to achieve the same quality in FoV.
- To achieve 4K resolution in FoV, a 360° video of *12K* x *6K* resolution is required.

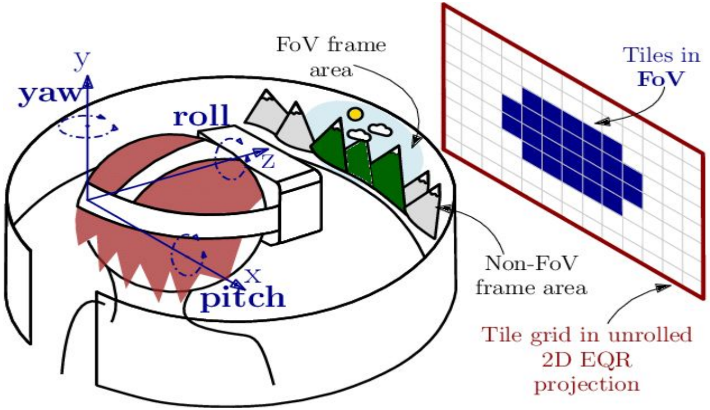


Fig. Users' Field of View (FoV)[2].

[1] Perfecto, Cristina et.al., "Taming the latency in multi-user VR 360°: A QoE-aware deep learning-aided multicast framework", TCOMM 2020.

Problem Statement

- User only sees FoV part of the frame, so fetching the complete spherical frame wastes the bandwidth.
- Existing methods (FoV-only[3], FoV+[4]) try to reduce bandwidth requirement by predicting user's FoV and streaming only FoV.
- Accurately predicting the user's FoV is challenging.
- The proposed approach tries to solve the issue of bandwidth usage without introducing any viewport prediction error.

[3] Qian, Feng et.al, "Optimizing 360 Video Delivery over Cellular Networks. Association for Computing Machinery", ATC 2016.
[4] Bao, Yanan et.al., "Motion-Prediction-Based Multicast for 360-Degree Video Transmissions", SECON 2017.

Existing Methods: FoV-Only

- It follows the well-known DASH protocol.
- It conducts the user trial to capture their head movements for different genres of the video.
- Linear regression is applied on the recorded user data to predict the viewport for short periods (e.g., 0.5s, 1s, 2s).
- It only fetches the tiles inside FoV, making it more susceptible to anomalies created by prediction error.

Fig. Tiled 360^{o} Video with highlighted FoV.

Existing Method: FoV+

- At any time instance *t* it predicts the viewport for *t+T* using collected movement traces of the user, where *T* denotes the prediction window.
- To deal with prediction error, it fetches redundant part of the frame which is situated outside the predicted FoV in low quality.
- It first predicts the planar FoV in the frame and then a circular part of that FoV is created which adds extra part of the frame around the edges of the planar FoV.

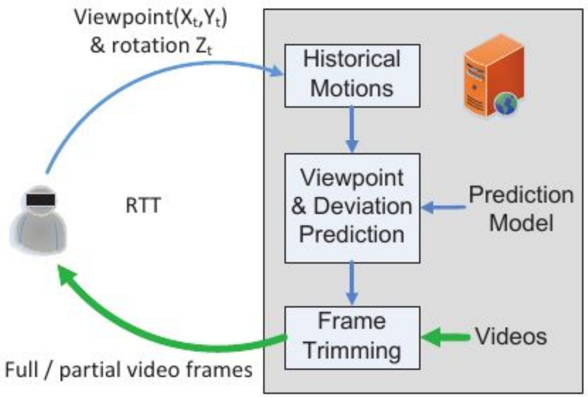


Fig. Working of FoV+[4].

[4] Bao, Yanan et.al., "Motion-Prediction-Based Multicast for 360-Degree Video Transmissions", SECON 2017.

Motivation : Saliency Map

- Saliency refers to unique features (pixels, resolution etc.) of the image in the context of visual processing.
- Saliency map is created by taking the average of feature maps (Colors, Intensity etc.).
- It helps in identifying important regions for proposed approach.

(a) Video frame

(b) Saliency Map

Fig. Saliency map of a snapshot from a soccer match.

Proposed Approach: Content Aware 360° Video Streaming

Saliency Map → Counting Salient Points → Threshold Selection → Saliency Count Lookup → Bitrate Allocation

Fig. Tiles of 360° video denoting normalized number of salient points.

Content Aware 360° Video Streaming: System Architecture

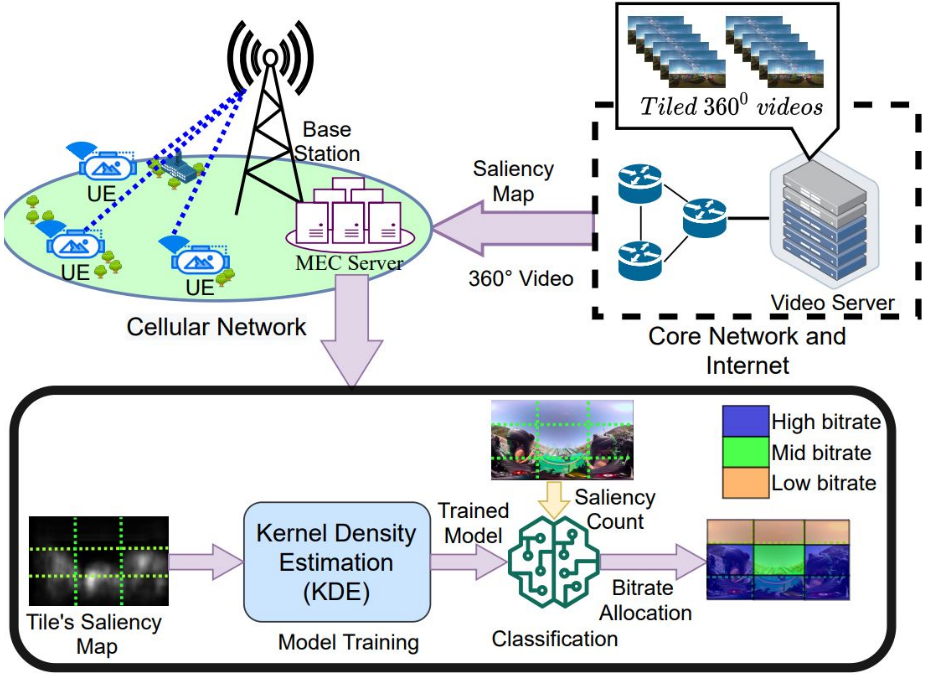


Fig. Architecture of the proposed content aware 360° video streaming framework.

Content Aware 360° Video Streaming: Threshold Selection

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

- Value of every pixel of saliency map ranges from 0 to 255.
- Count of salient points in one tile can range have a wide range.
- Saliency counts are not evenly distributed.
- Classification of saliency counts is necessary for bitrate allocation.
- Saliency counts have a many-to-one relationship with available bitrates.

5	210	100	175
190	15	98	185
200	150	65	53
204	190	35	47

Fig. Representation of saliency lookup.

Content Aware 360° Video Streaming: Bitrate Allocation

- The proposed method doesn’t differentiate between the FoV region and the OOS region.
- All tiles in FoV do not necessarily contain the salient regions.

Fig. Tiled video with highlighted FoV.

Results: 360° Video Dataset[1]

- Dataset is derived from 10 standard 360 videos from YouTube with different characteristics.
- 60 seconds parts of the video are extracted from every video.
- It uses convolutional neural network for saliency map extraction.

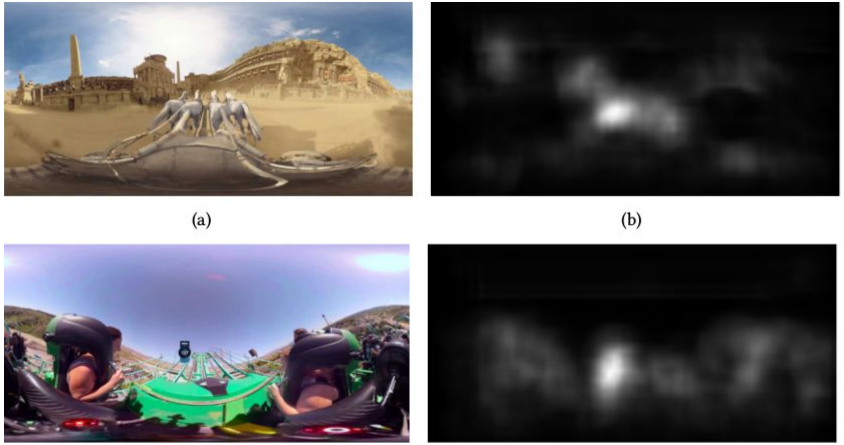
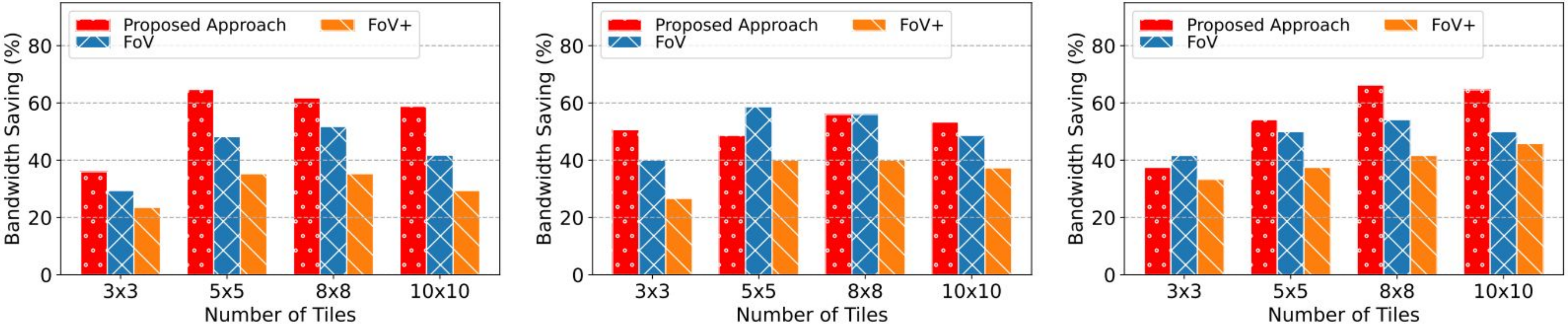


Fig. Illustration of saliency maps of two 360° video frames[1].

[1] Wen-Chih Lo et.al., “360° Video Viewing Dataset in Head-Mounted Virtual Reality,” MMSys'17

Results

- Distribution of salient regions directly affects the bandwidth saving.
- Driving 360 has sparsely distributed salient objects, whereas Kangaroo Island has dense distribution over a large area of the frame.

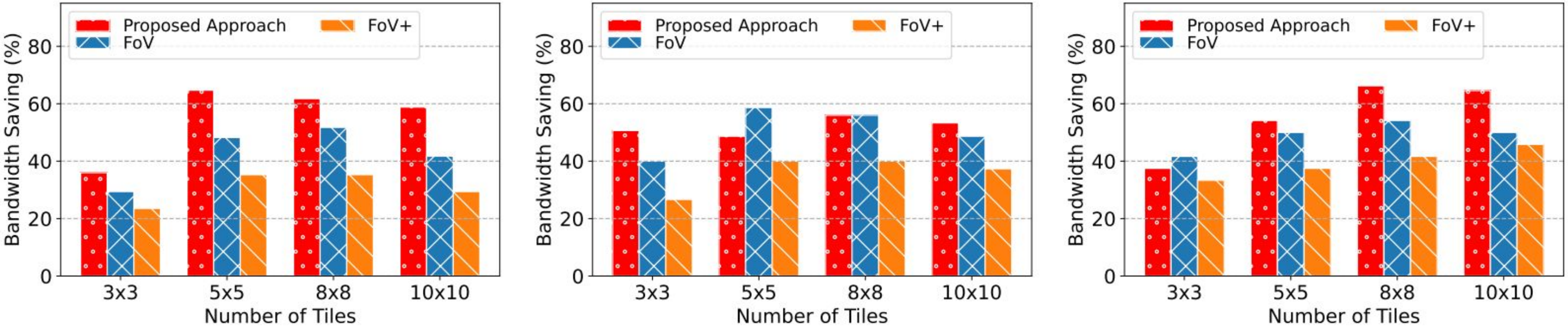


(a) Driving 360° (b) Kangaroo Island (c) Hog Ride

Fig. Bandwidth Saving Comparison for the different videos from the dataset.

Results

- Dense salient object distribution over very small part of the frame in Roller Coaster and Shark Dive Video.



(a) Driving 360° (b) Kangaroo Island (c) Hog Ride

Fig. Bandwidth Saving for the different videos from the dataset.

Results

- Effect of tile variation on bandwidth reduction is different for different videos.

Video	3x3	5x5	8x8	10x10	Original
Driving 360	11 MB	6 MB	6.5 MB	7 MB	17 MB
Kangaroo Island	7.4 MB	7.7 MB	6.6 MB	7 MB	15 MB
Pacman	4.3 MB	4.5 MB	5.4 MB	6.5 MB	10 MB
Hog Rider	15 MB	11 MB	8.1 MB	8.5 MB	24 MB
Roller Coaster	7.7 MB	4 MB	4.9 MB	6.1 MB	16 MB

Fig. Bandwidth reduction of different types of 360° videos for different variations of tiling.

Results

- QA of video created with proposed approach outperforms FoV with huge margin.
- QA of video created with proposed approach outperforms FoV+ without having to predict future FoV.

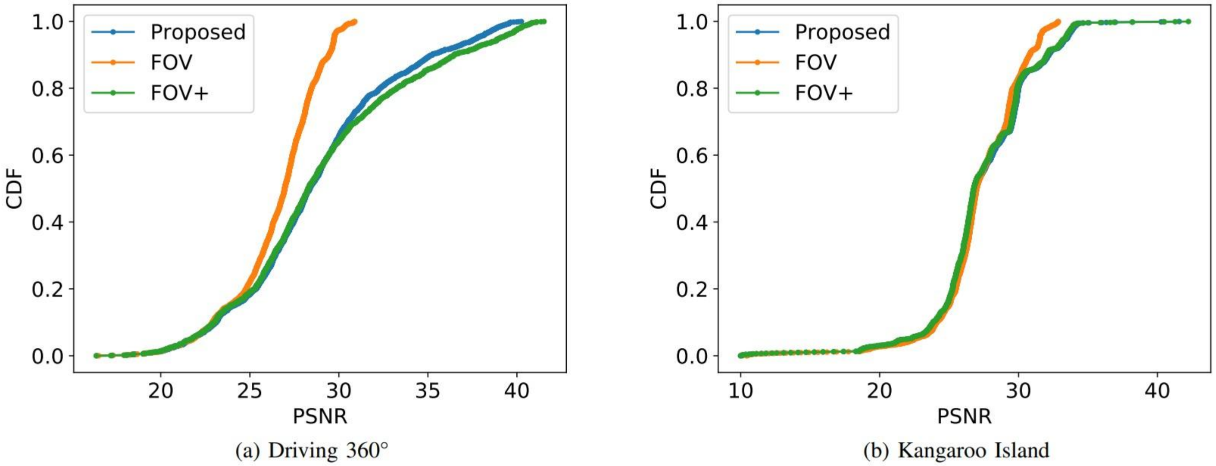


Fig. QA achieved using the proposed solution for the videos in the dataset.

Conclusion

- This work introduces a content-aware adaptive bitrate selection solution for 360° video streaming.
- The proposed solution determines the important video tiles using the saliency map which helps to overcome the dependency on the client feedback for the user’s viewport.
- The results from our extensive experiments show that the proposed solution achieves up to 50% bandwidth reduction as compared to the legacy 360° video streaming.
- The proposed solution saves 15% more bandwidth while maintaining an average 28.5 dB PSNR as compared to the existing solutions.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Thank You !

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

cs15resch11011@iith.ac.in

Contact us for Queries.